2010 11th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing

Bagging of Complementary Neural Networks with Double Dynamic Weight Averaging

Sathit Nakkrasae*, Pawalai Kraipeerapun*, Somkid Amornsamankul† and Chun Che Fung‡

*Department of Computer Science, Faculty of Science
Ramkhamhaeng University, Bangkok, Thailand
Email:sathit@ru.ac.th, pawalai@ru.ac.th

†Department of Mathematics, Faculty of Science
Mahidol University Bangkok,Thailand and
Centre of Excellence in Mathematics CHE
Sriayudhaya Rd., Bangkok, Thailand
Email:scsam@mahidol.ac.th

‡School of Information Technology
Murdoch University, Perth, Australia
Email:l.fung@murdoch.edu.au

***Abstract*—Ensemble technique has been widely applied in regression problems. This paper proposes a novel approach of the ensemble of Complementary Neural Network (CMTNN) using double dynamic weight averaging. In order to enhance the diversity in the ensemble, different training datasets created based on bagging technique are applied to an ensemble of pairs of feed-forward back-propagation neural networks created to predict the level of truth and falsity values. In order to obtain more accuracy, uncertainties in the prediction of truth and falsity values are used to weight the prediction results in two steps. In the first step, the weight is used to average the truth and the falsity values whereas the weight in the second step is used to calculate the final regression output. The proposed approach has been tested with benchmarking UCI data sets. The results derived from our technique improve the prediction performance while compared to the traditional ensemble of neural networks which is predicted based on only the truth values. Furthermore, the obtained results from our novel approach outperform the results from the existing ensemble of complementary neural network.**

***Keywords*-Backpropagation Neural Network; Complementary Neural Networks; Diversity; Bagging; Ensemble;**

I. Introduction

In general, applying a single system to solve the large and complex real world problems might not be adequate for good performance [4]. Therefore, several researchers [5], [6], [7], [12], [13], [14], [15] have utilized the ensemble system to improve the accuracy on their works. One aspect that can improve the prediction performance of the ensemble technique is the system diversity [3]. Diversity is a measure that defines the disagreement degree in the output of the individual classified machines in the ensemble [11]. All machines in the ensemble should be diverse among themselves. Beside the diversity, accuracy of individual classifiers is also an important consideration. To derive better performance of ensemble, the trade-off between diversity and accuracy have to be considered [3]. It was found that a diverse ensemble of less accurate classification can yield better performance than an ensemble of more accurate classification with less diversity [17].

Diversity in an ensemble system can be managed when the individual classified machines is created under different situations, which are different parameter settings of the classifiers, different classifier training datasets, and different classifier types [3]. In this paper, diversity can be reached by creating the individual classified machine under different classifier training datasets using bagging learning strategies with N components in an ensemble [2]. Bagging technique provides N views of the original training set which are generated by sampling with replacement procedure. All outputs obtained from all machines in the ensemble are aggregated in order to compute the final prediction output as shown in Fig. 1

In [8], an ensemble of complementary neural networks (CMTNN) was utilized to solve the binary classification problem. The result was shown that the ensemble of CMTNN can yield better performance than a single system of CMTNN. CMTNN consists of pairs of neural networks which are the truth neural network trained to predict the level of truth values and the falsity neural network trained to predict the level of falsity values. The falsity value is supposed to be complement to the truth value of each input pattern, however, the boundary between the predicted truth and the predicted falsity values can be overlapped. Therefore, uncertainty can occur in this situation. This uncertainty value is computed from the difference between the truth and the falsity value which is used to calculate the weight in order to enhance the predicted result.

In this study, an ensemble of CMTNN is used to solved the regression problem. Uncertainty values are also applied

978-0-7695-4088-7/10 $26.00 © 2010 IEEE
DOI 10.1109/SNPD.2010.34

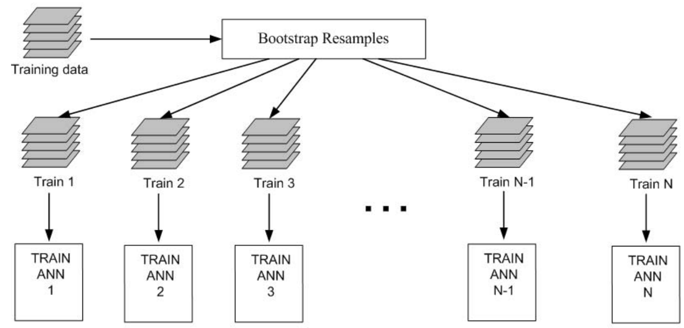

Figure 1. The traditional ensemble technique based on bagging approach

to provide better output result. Uncertainty is computed from the prediction of the truth and falsity values which are then used to compute the weight. The proposed aggregation technique is separated into two parts. In the first part, uncertainty values are used to calculate the dynamic weight which is used to enhance the average truth and and the average falsity values. In the second part, uncertainty values are used to compute the dynamic weight for the averaged truth and the averaged falsity obtained from the first step. The benchmark data from UCI [1] are used to test our novel approach. These data sets are computer hardware, concrete compressive strength, and housing. These classical data sets were also used in [9], [10], [16]. The prediction performance of our method is compared with the existing techniques such as the ensemble of feed forward back propagation neural networks (BPNN), the ensemble of complementary neural networks (CMTNN) based on equal weight averaging, and the ensemble of complementary neural networks (CMTNN) based on dynamic weight averaging.

The rest of this paper is organized as follow. Section II explains an ensemble of complementary neural networks with the proposed aggregation technique. Section III describes the data sets and the experiment results. Conclusion and future work are described in section IV.

II. Ensemble of CMTNN with double dynamic weight averaging

In this paper, neural networks in the training step are diverse and all the obtained outputs are aggregated. Diversity is managed using bagging technique in cooperated with bootstrap re-sampling in order to generate m training sets for m components in the ensemble. Each generated training set is created by random selection with replacement of the original input patterns. Instead of using native back propagation neural network in each component, complementary neural networks are used. Each component in the ensemble is composed of a pair of neural networks named the truth and the falsity neural network. Both networks apply the opposite target values. The truth neural network is trained to predict the level of the truth values while the falsity neural network is trained to predict the level of the falsity values as showed in Fig. 2.

Although these two neural networks are used to train different target values, the number of neurons, initial weight and parameter setting are the same. The same generated training data are used in both truth and falsity neural networks in each component. The only different fashion is the falsity neural network applied the complement target value of the truth neural network. For example if the truth neural network applied 0.8 as the target value then the target value of the same input pattern for the falsity neural network will be 0.2.

In this paper, double dynamic weight averaging technique is used to aggregate the outputs obtained from all components in the ensemble. These weights are considered as certainty in the prediction. The high certainty means the high weight. Fig. 3 portrays the proposed ensemble of CMTNN with double dynamic weight averaging in the testing phase. Double dynamic weight in this study means that the dynamic weights are used in two steps. In the first step, uncertainty is considered as the difference between the truth and the non-falsity output for each pattern in each component. Hence, the weight is calculated from the uncertainty value and then used to weight the averaged truth and the averaged falsity values.

Let $T_j(x_i)$ and $F_j(x_i)$ be the truth and the falsity values predicted for the input pattern x_i of the component j, where $j = 1, 2, 3, , m$. The uncertainty values can be calculated as follows:

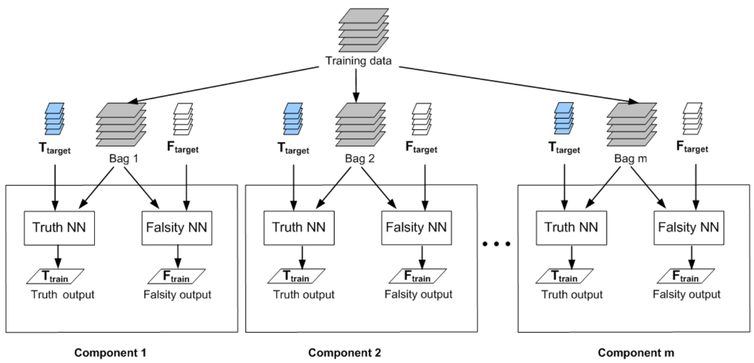

Figure 2. The ensemble of complementary neural networks based on bagging technique. (training phase)

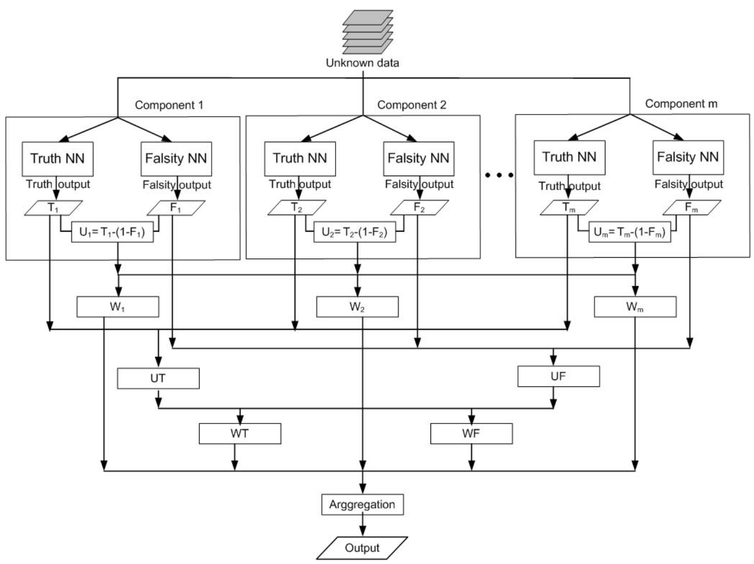

Figure 3. The proposed ensemble of complementary neural networks with double dynamic weight averaging. (testing phase)

$$U_j(x_i) = T_j(x_i) - (1 - F_j(x_i)) \tag{1}$$

Let $W_j(x_i)$ be the weight derived from the j^{th} component for the input pattern x_i which can be computed as follows:

$$W_j(x_i) = \frac{1 - U_j(x_i)}{\sum_{j=1}^{m} (1 - U_j(x_i))} \tag{2}$$

Let $T_{weight}(x_i)$ and $F_{weight}(x_i)$ be the weighted average truth output and the weighted average falsity output, respectively. Both values can be calculated as follows:

$$T_{weight}(x_i) = \sum\nolimits_{j=1}^{m} (W_j(x_i) \times T_j(x_i)) \tag{3}$$

$$F_{weight}(x_i) = \sum\nolimits_{j=1}^{m} (W_j(x_i) \times F_j(x_i)) \tag{4}$$

For each input pattern, m components in the ensemble provide m different truth and falsity values. All the truth values are not exactly the same value. Also, all the falsity values are not the same. Therefore, uncertainly value obtained from these difference is used to improve the aggregate regression output. In the second step, The average of the difference among the entire truth values for each input pattern is calculated. In the same manner, the average of the difference among the entire falsity values of each input pattern is computed. Let $U_T(x_i)$ be an uncertainty obtained from the difference among the truth values of the input pattern x_i. Let $U_F(x_i)$ be an uncertainty obtained from the difference among falsity values of the input pattern x_i. $U_T(x_i)$ and $U_F(x_i)$ can be computed as follows:

$$U_T(x_i) = \frac{\sum_{k,h=1}^{m} |T_k(x_i) - T_h(x_i)|}{m(m-1)/2}; k \neq h \tag{5}$$

$$U_F(x_i) = \frac{\sum_{k,h=1}^{m} |F_k(x_i) - F_h(x_i)|}{m(m-1)/2}; k \neq h \tag{6}$$

Both uncertainly values are used to weight the combination between the weighted average truth value and the weighted average falsity value obtained from the first step. The weight for the truth value is calculated as the complement of the $U_T(x_i)$ while the weight for the falsity value is computed as the complement of the $U_F(x_i)$.

Let $W_T(x_i)$ and $W_F(x_i)$ be the weight for the truth value and the weight for the falsity value, respectively. The regression output with double dynamic weight averaging $O(x_i)$) can be computed as follow:

$$O(x_i) = \begin{array}{c}(W_T(x_i) \times T_{weight}(x_i))+ \\ (W_F(x_i) \times (1 - F_{weight}(x_i)))\end{array} \tag{7}$$

$$W_T(x_i) = \frac{1 - U_T(x_i)}{(1 - U_T(x_i)) + (1 - U_F(x_i))} \tag{8}$$

$$W_F(x_i) = \frac{1 - U_F(x_i)}{(1 - U_T(x_i)) + (1 - U_F(x_i))} \tag{9}$$

III. Experiments

A. Data Sets Preparation

Benchmark data sets named computer hardware, concrete compression strength, and housing from UCI repository are used in this experiment. Table I show the characteristic of these data sets. Each data set is randomly split into 80% training set and 20% testing set.

B. Experimental Procedure, Results And Analysis

Three UCI data sets are used to test the proposed ensemble model. This experiment does not focus on the optimization of the prediction but the purpose is only on the improvement of the prediction. For each data set, we create four types of ensemble, which are the proposed ensemble of complementary neural networks with double dynamic weight averaging, the ensemble of complementary neural networks with equal weight averaging, the ensemble of complementary neural networks with dynamic weight averaging, and the ensemble of feed forward back propagation neural network (BPNN). Each neural network in CMTNN is applied based on feed forward back propagation neural network. The obtained results from these four types of ensemble will be compared. All neural networks in this experiment are constructed using the same architecture and parameters. Thirty bags of training data sets are created using bootstrap resampling with replacement and applied to thirty components in each ensemble. For each neural network, the number of input-nodes is equal to the input features which are 6, 8, and 13 for computer hardware, concrete, and housing data sets, respectively. They have one hidden layer constituting of $2n$ neurons where n is the number of input features.

With the purpose of output aggregation, outputs from all components in the ensemble of BPNN are averaged. For the ensemble CMTNN, the aggregation techniques which are equal weight averaging and dynamic weight averaging are applied. For our novel aggregation approach, double dynamic weight averaging is applied to CMTNN.

In CMTNN technique, the truth and falsity outputs are created from a pair of the truth and the falsity neural networks in each component. For the equal weight averaging technique applied to CMTNN, a simple averaging is applied to these truth and non-falsity outputs. For the dynamic weight averaging technique, the weight is calculated only based on equations (8) and (9). This technique does not consider the weight computed from equation (2). The average truth and the average falsity are computed from the simple averaging technique and then used to calculate the final output.

For our proposed aggregation technique, uncertainty values are used to compute double dynamic weight averaging. In the first part, the weight is computed from the uncertainty value which is considered as the difference between the truth and the non-falsity output in each component. In the second part, the weights are calculated based on the difference among the truth values and the difference among the falsity values from all components. The weight obtained from the first part is used for the averaged truth and the averaged falsity values while the weight obtained from the second part is used to calculate the final regression output.

Table II shows mean square error (MSE) derived from our novel ensemble of CMTNN compared to the existing ensemble of CMTNN and BPNN. This table shows that our proposed ensemble of CMTNN with double dynamic weight averaging yield better performance than other techniques. Table III shows the percent improvement of our purposed ensemble CMTNN with double dynamic weight averaging compared to other existing techniques. According to the results obtained from Concrete, Hardware, and Housing data sets, it can be shown that our purposed technique compared to BPNN provides better performance with 39.83%, 57.77%, and 8.25% respectively. For CMTNN with equal weight averaging, our purposed approach yield better performance with 6.80%, 2.40%, and 4.71% respectively. Our proposed technique also gives superior results with 1.42%, 0.49% and 0.8% when compared to CMTNN with dynamic weight averaging.

IV. Conclusion and Future Work

The aim of this paper is trying to adjust and improve the existing CMTNN approach. An ensemble of CMTNN with double dynamic weight averaging based on bagging technique has been constructed to solve the regression problem. Each CMTNN consists of a pair of neural networks which are used to predict the truth and the falsity values. Uncertainty in the prediction has been used to enhance the prediction performance. There are two steps to apply the weight in our purposed aggregation technique. The weight derived from the first step is used to average the truth value and the falsity value for the input pattern while the weight in the second step is utilized to calculate the final regression output. The result shows that our proposed approach yield

Table I
DATA SETS FROM UCI REPOSITORY.

Name	Feature type	No. of features	No. of samples	No. of training data	No. of testing data
Concrete	numeric	8	1030	824	206
Hardware	numeric	6	209	167	42
Housing	numeric	13	506	405	101

Table II
THE COMPARISON AMONG THE MEAN SQUARE ERROR (MSE) OBTAINED FROM THE ENSEMBLE OF BPNN, CMTNN WITH EQUAL WEIGHTED AVERAGING, CMTNN WITH DYNAMIC WEIGHT AVERAGING AND CMTNN WITH DOUBLE DYNAMIC WEIGHT AVERAGING FOR CONCRETE, HARDWARE, AND HOUSING DATA SETS.

Ensemble of	Concrete	Hardware	Housing
BPNN	0.019303	0.004419	0.008398
CMTNN with equal weight averaging	0.012462	0.001912	0.008087
CMTNN with dynamic weight averaging	0.011783	0.001875	0.007773
CMTNN with double dynamic weight averaging	0.011615	0.001866	0.007705

Table III
THE PERCENT IMPROVEMENT OF THE PROPOSED ENSEMBLE OF CMTNN WITH DOUBLE WEIGHT AVERAGING COMPARED TO THE ENSEMBLE OF BPNN, CMTNN WITH EQUAL WEIGHTED AVERAGING AND CMTNN WITH DYNAMIC WEIGHT AVERAGING.

Ensemble of	Concrete	Hardware	Housing
BPNN	39.83%	57.77%	8.25%
CMTNN with equal weight averaging	6.80%	2.40%	4.71%
CMTNN with dynamic weight averaging	1.42%	0.49%	0.87%

better accuracy than BPNN ensemble, CMTNN ensemble with equal weight averaging and CMTNN ensemble with dynamic weight averaging. Although our purposed ensemble CMTNN with double dynamic weight averaging technique provides minor improvement when compared to both existing CMTNN techniques, it is obvious to see that the CMTNN is a superior technique when compared to other existing techniques. In our next study, other ensemble techniques and other types of uncertainty will be considered and used to solve the regression problem.

REFERENCES

[1] A. Asuncion and D. Newman, "UCI Machine Learning Repository," http://www.ics.uci.edu/∼mlearn/MLRepository.html, 2007, university of California, Irvine, School of Information and Computer Sciences.

[2] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[3] D. F. De Oliveira, A. M. P. Canuto, and M. C. P. De Souto, "The diversity/accuracy dilemma: an empirical analysis in the context of heterogeneous ensembles," in *CEC'09: Proceedings of the Eleventh conference on Congress on Evolutionary Computation*. Institute of Electrical and Electronics Engineers Inc., The, 2009, pp. 939–946.

[4] L. Hansen and P. Salamon, "Neural network ensembles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 10, pp. 993–1001, 1990.

[5] R. Ismail and O. Mutanga, "A comparison of regression tree ensembles: Predicting sirex noctilio induced water stress in pinus patula forests of kwazulu-natal, south africa," *International Journal of Applied Earth Observation and Geoinformation*, vol. In Press, Corrected Proof, pp. –, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/B6X2F-4XF7XYP-1/2/c4c58b187faed4f7c261ab9dcaaa6563

[6] U. Johansson, T. Lofstrom, and L. Niklasson, "Obtaining accurate neural network ensembles," *Computational Intelligence for Modelling, Control and Automation, International Conference on*, vol. 2, pp. 103–108, 2005.

[7] M.-J. Kim and D.-K. Kang, "Ensemble with neural networks for bankruptcy prediction," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3373 – 3379, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/B6V03-4XFY0F0-K/2/799fd01c273fb0584cccba25ff212f86

[8] P. Kraipeerapun and C. C. Fung, "Binary classification using ensemble neural networks and interval neutrosophic sets," *Neurocomputing*, vol. 72, no. 13-15, pp. 2845–2856, 2009.

[9] P. Kraipeerapun, S. Nakkrasae, S. Amornsamankul, and C. C. Fung, "Complementary neural networks for regression problems," in *ICMLC '09: 2009 International Conference on Machine Learning and Cybernetics*, 2009, pp. 3442–3447.

[10] G. Liu, J. Pang, X. Piao, and S. Huang, "The discovery of attribute feature cluster for any clustering result based on outlier detection technique," in *ICICSE '08: Proceedings of the 2008 International Conference on Internet Computing in Science and Engineering*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 68–72.

[11] P. Melville and R. J. Mooney, "Constructing diverse classifier ensembles using artificial training examples," in *IJCAI*, 2003, pp. 505–512.

[12] Y. L. Murphey, Z. Chen, M. Abou-Nasr, R. Baker, T. Feldkamp, and I. Kolmanovsky, "Ensembles of neural networks with generalization capabilities for vehicle fault diagnostics," *Neural Networks, IEEE - INNS - ENNS International Joint Conference on*, vol. 0, pp. 2188–2194, 2009.

[13] C.-F. Tsai and J.-W. Wu, "Using neural network ensembles for bankruptcy prediction and credit scoring," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2639 – 2649, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/B6V03-4NRT397-1/2/655521c2483a82dd2ca8eb06d6c46e42

[14] W. Wang, D. Partridge, and J. Etherington, "Hybrid ensembles and coincident-failure diversity," in *IJCNN '01 : 2001 International Joint Conference on Neural Networks*, 2001, pp. 2376–2381.

[15] J. Wu, "A novel artificial neural network ensemble model based on k–nearest neighbor nonparametric estimation of regression function and its application for rainfall forecasting," in *CSO (2)*, 2009, pp. 44–48.

[16] I.-C. Yeh, "Modeling of strength of high performance concrete using artificial neural networks," *Cement and Concrete Research*, vol. 28, no. 12, pp. 1797–1808, 1998.

[17] G. Zenobi and P. Cunningham, "Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error," in *ECML*, 2001, pp. 576–587.